ESTIMATION OF NUTRITIONAL VALUES OF FOOD USING INCEPTION v3

Yash Gosalia[1], Meet Karnik[2], Aayush Pandey[3]

[1,2,3]*Dept. of Information Technology, Atharva College of Engineering, Mumbai.*

Abstract - *Today's generation is very aware of what they are eating and the amount of calories in their food. Eating too many calories can lead to increased weight, which has become a big health issue. A study from 2016 states that more than 1.9 billion adults are overweight where almost one third of these are obese. Measuring food calorie intake is one of the important steps towards fighting obesity. One of the methods includes manually tracking our food and calorie intake but that method would be tedious for many users. Maintaining a balanced diet is important for living a healthy life. There has been a rapid increase in diseases mainly due to an inconsistent diet and too much consumption of calories by users. Systems that can identify food items based on real time images and analyze its nutritional values can be beneficial for users to achieve a more balanced diet This paper proposes a system to automatically estimate nutritional values such as proteins, carbohydrates, minerals, etc. by classifying the input image of food. Our method employs a deep learning model for accurate food identification. There are several steps taken to recognize food images, which are to collect data using libraries in python, text processing, testing training data, and text classification using the Support Vector Machine method. The Support Vector Machine method is used to help classify classes of food. In addition to image analysis, food attributes are estimated by CalorieNinjas API which would provide the nutritional values required by the user. The combined data was tested from the training data used for each food image and got an accuracy of 85%.*

***Key Words*: Food Recognition, Support Vector Machine, Deep Learning, Text Mining, Convolutional Neural Networks.**

1. INTRODUCTION

Food is an essential part of everyday life. What we eat can have a big impact on our weight and according to the World Health Organization (WHO) the overweight and obesity has risen drastically in the last 40 years. [6] In 2016 more than 1.9 billion adults worldwide were overweight, where over 650 million of these were obese. If we consider the statistics of Norway for the year 2017, Norwegian Institute of Public Health (NIPH) stated in a report last updated late 2017 that 1 out of 4 men and 1 out of 5 women are obese. This is a huge concern and is also a big risk factor for diseases such as diabetes, heart disease, musculoskeletal disorders and some different types of cancers.

Maintaining a healthy and balanced diet is a concern for many individuals. One way of achieving this is by keeping a track of the consumed calories. This method can be very tiring as it requires the user to do a lot of manual entries and calculations to obtain the result. In fact, it has been also shown that people tend to underestimate the number of calories they are consuming most of the time. Recently, automated methods of determining calorie content of food have been developed. For instance, many applications exist nowadays to enable users to calculate the values. Most of these tools, however, require the user to enter some information about the food item consumed. For instance, it expects that the user will enter the name of the food item or the ingredients, as well as the size of the food item. These tools typically require input from the user which is then run against a database of the food items to do the necessary calculations. [1] In this paper, we propose to alleviate the user from the burden of entering the above information in order to calculate the number of calories consumed in a food item. Our proposed approach works as follows.

The proposed system aims to be a step towards creating awareness based on health and fitness concerns so that people can eat and live in a better way. The proposed method helps in determining the nutritional content of food automatically by making it feasible for a person to learn about what food might contain and how healthy it might be. [3] The inherent theme is to automatically detect food items from an image of a platter and then estimate the respective food attributes such as the percentage of calcium, iron etc. along with the ingredients present in the food. Our system provides nutritional values as found in commercially available food items. The proposed system has its application in health-care industries and hospitals. Knowing about the nutritional value will further provide motivation for patients to refrain from food that can be detrimental to their health.

The user provides the system with an image of the food item. Based on the features of the image and a supervised machine-learning model, the system identifies the type of food item in the image It also predicts the serving size in grams and then using these two obtained results and the original submitted image, the model predicts an estimate of the nutrients present in the food item.. We show that with the help of this method we can estimate the calories and the nutritional content of the food item with accuracy. Even though a few methods to achieve the same result do exist, none of them have been shown to operate with such accuracy. Moreover, these exiting systems had a very limited applicability i.e. they would not offer wide varieties of food items and would in most cases require additional information. We propose to build a highly accurate machine learning model to obtain the desired results. We also aim to provide training datasets that would be used by us and can also be used for further research and development in this field.

2. RELATED WORK

Manal Chokr & Shady Elbassouni [2] proposed a system in which the user submits an image of the food item to the system. Based on the image visual features and by means of a supervised machine-learning model, their system is able to predict what the type of the food item is. Based on the dimensions predicted by the system and the type of food item, numbers of calories in a food item are predicted. But this system can only predict the number of calories in the food item.

Hammad Afzal et al. [1] proposed a system which consists of two components. The primary element uses CNNs to acknowledge the food item in a picture. The second element estimates food attributes mistreatment text retrieval from net archives further as scrapping of information from biological process and instruction websites for ingredients and nutrient counts. This data is trained on a two layer neural network, from which we can compute probabilities of existing ingredients in a particular food item. Each food item typically has a standard serving size against which calories and nutritional content can be calculated. The system uses deep learning algorithms, a server with a trained model to recognize food images and estimate its attributes along with ingredients, and a conventional mobile application. The classifier requires a large dataset containing multiple images against every category of food item for training purposes.

Pouladzadeh et al. [5] planned a system that involves capturing an image of the food and processing it through predefined steps, which follow pipeline architecture. These steps include food image segmentation and food portion recognition. Calorie measurement is done using nutritional fact tables. The system often fails to detect various food portions in mixed food; it also fails to segment them properly. The Area measurement technique proposed is based on a depth estimation technique. However, their system uses a dataset that is too simplistic, consisting of food items placed on whiteplates with smooth texture.

Taichi Jontou et al. [3] introduced a MKL-based feature fusion method into food image recognition. In their recognition, we prepare 50 kinds of food categories as shown in and classify an unknown food image into one of the categories. In the training step, Image features such as bag-of-features (BoF), color histogram and Gabor texture features are extracted from the training images, and we train a MKL-SVM with extracted features. The same steps are followed for classifying an unknown image into the given category. The system has a limited dataset and is proven to be effective for a limited number of categories.

Vinay Bettadapura et al[4] proposed a system which consists of An automatic workflow where online resources are queried with contextual sensor data to find food images and additional information such as geographic location, name of restaurant, etc. so that food can be recognized more efficiently. An image classification approach using the SMO-MKL multi-class SVM classification framework with features extracted from test photographs. This system is completely dependent on additional data and can't classify food items based on the images alone.

3. METHODOLOGY

The proposed system contains two major modules:

• Food Recognition:

Recognizing food items from images

• Nutritional Value Estimation:

Estimating nutritional value of the recognized food item using API.

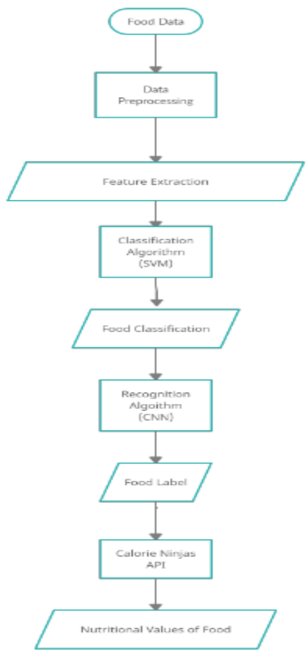

Fig-1: Flow Diagram of the Proposed Model.

Fig.1 shows the flow of the model where images are represented in the form of pixels. Classification is performed using the linear kernel SVM (support vector machine). The position of the camera and viewing direction has to be constant in order to obtain more efficient classifications. Convolutional neural networks have also been employed for the recognition task and as a result the recognition accuracy has improved significantly.

For food attribute estimation

For attribute estimation, the image of the food item after being identified is queried against an API containing nutritional data such as vitamins, minerals, fats, etc. For over 100000 foods and beverages as well.

3.1 Dataset Description

Our goal is to make a dataset that contains common food items, augmented with sub continental dishes. We started by experimenting on the publicly available dataset of food images, i.e. Food-101. It contains 101 classes of food items with 1000 images for each class. Food-101 is designed specifically for classifying multiple foods. [7] There are other datasets that have been used for food recognition previously; one such dataset is Food-5k. Food-5k contains 5000 images, out of which 2500 are of food and 2500 of non-food. However, this dataset can be used only for binary classification to discriminate food items from non-food items and therefore, is not suitable for our task. Moreover, Food-101 does not include food items or classes from the sub continental cuisine which makes a large portion of the food that people intake in the sub continental region. Some sub continental dishes exhibit low inter-class variation and are very similar to each other, so collecting high quality data for accurate classification of different categories is a big challenge. The results returned by Google search engine against textual search queries for food images are quite relevant with very low noise content. Based on such results from Google search engine, our new dataset is created by querying Google against each label of our dataset. The newly formed dataset consists of foods that are common in almost every country. Thus, the final dataset contains all the food items from the Food-101 dataset and 100 additional sub continental food classes. The dataset is split into training and validation images. Each class contains around 800 training images and 200 validation images.

Fig-2: Food 101 Dataset

3.2 Data Preprocessing

The dataset contains food images. The Food-101 dataset consists of one folder containing all the categories. In order to use the dataset, there has to be some testing data. This has been done by taking 100 random images from each category and creating testing data from them. The dataset will then have a training set of 99 990 images and a testing set of 1100 images.[8] The main goal is to automatically recognize food items from the user when he captures it from his camera. On purpose, the training images were not cleaned of some noise which is in the form of intense colors and wrong labels. All images were resized to a maximum length of 512 pixels. This task was done using an Inception pre-trained model. Using these pre-trained models, we can use the already learned weights and add a few layers on top to fine tune the model to our new data. This helps in faster convergence and saves time and computation when compared to models trained from scratch.

3.3 Model Architecture and Implementation

Convolutional neural networks are a variation of the better-known Multilayer Perceptron in which node connections are inspired by the visual cortex. CNNs have proven to be very beneficial in image recognition, video analysis and natural language processing.

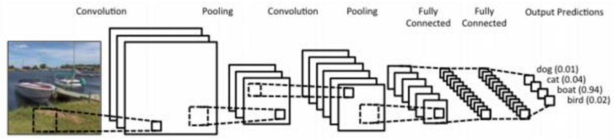

Fig-3: CNN Model Architecture

CNNs take images as input. A filter (kernel) is subsequently slid over the food image and patterns for food-item and non-food items are learned. CNN usually comprises of several pairs of convolution and max-pooling layers. Two significant ideas make the CNN more useful and effective, i.e. local connectivity and weight sharing. Local connectivity restricts that each neuron connects to only a local region of inputs, leading to sparse interactions; and weight sharing reduces the number of the parameters and makes CNN much more reliable.

The CNN begins by learning features like vertical lines, but in subsequent layers, begins to pick up on features like the shape of food-item. Such learned features may provide an elegant and powerful representation of different prosodic features of image, which in turn are representative of underlying differences between food-item and non-food items. However, with the highly detailed representations, false images can be inconveniently picked up by the network. One can mitigate this noise with different regularization parameters in the network (pooling layers, L1 loss functions, dropout, etc.),

For our proposed system, we plan to use a pre-trained model called Inception V3.Inception-v3 is a convolutional neural network architecture from the Inception family that makes several improvements including using Label Smoothing, Factorized 7 x 7 convolutions, and the use of an auxiliary classifier to propagate label information lower down the network (along with the use of batch normalization for layers in the side head)

Inception v3 mainly focuses on burning less computational power by modifying the previous Inception architectures. The architecture of an Inception v3 network is progressively built, step-by-step, as explained below:

1. Factorized Convolutions: This step helps to reduce the number of input characteristics required in a network and keeps checking the network efficiency.

2. Smaller convolutions: Instead of bigger convolutions, smaller convolutions are used to decrease training time. Say a 6 × 6 filter has 36 parameters; two 4 × 4 filters replacing a 6 × 6 convolution has only 32 (4*4 + 4*4) parameters instead.

3. Asymmetric convolutions: A three × three convolutions could be replaced by a one × three convolution followed by a three × one convolution. If a three × three convolution is replaced by a two × two convolution, the quantity of parameters would be slightly more than the asymmetric convolution proposed.

4. Auxiliary classifier: It is a tiny CNN inserted between layers during training, and the loss incurred in that network is added to the overall network loss. In GoogLeNet auxiliary classifiers were used to increase the number of layers in a network, whereas in Inception v3 an auxiliary classifier acts as a regularizer.

5. Grid size reduction: This is usually done by pooling operations. To decrease computational cost, a more efficient technique is proposed

3.4 Training the Model

This work utilizes the model of Google Inception-V3 that is pre-trained on the Food-101, where the reshaped size ―299x 299 x 3" is considered for all the images. Moreover, the average-pooling function is considered on the food image dataset and takes average of all the image features. The space output dimensionality is defined through the dense-function. However, 0.5 dropout fraction-rate of input units is taken to overcome the issue of over-fitting. Additionally, to decide the definite class from the several numbers of classes, the function of softmax is defined and it identifies the maximum probability in order to obtain output for the particular class and neglect the rest of the classes. Here, a CNN is used to get effective food image classification, the stochastic Gradient- Descent is considered with a rapidly reducing learning to obtain a better performance. The 30 epochs has been considered to train a model and have defined callbacks to record the growth using a log file. Moreover, a learning rate of a scheduler is defined, where it takes input of epoch index and afterwards provides an output of new learning rate. Check pointer callback is used to build the model checkpoints and these are saved in the format of .hdf5 files, where only the best score is considered to save the learned models.

4. ANALYSIS AND RESULTS

In the early stages of our work, we were able to successfully classify 3 classes of food items, which gave our model an initial accuracy rate of 70% and 30% loss rate for 20 epochs with a batch size of 16. The reason for the high loss rate is the images varied in sizes and pixels, some low-light photos. In the later stages, we were able to successfully identify 11 classes of food items, which gave our model an accuracy rate of 75% and 25% loss rate for 25 epochs with a batch size of 16.

Finally, we were able to successfully identify 50 classes of food items, which have our model an accuracy rate of 80% and 20% loss rate for 30 epochs with a batch size of 16.

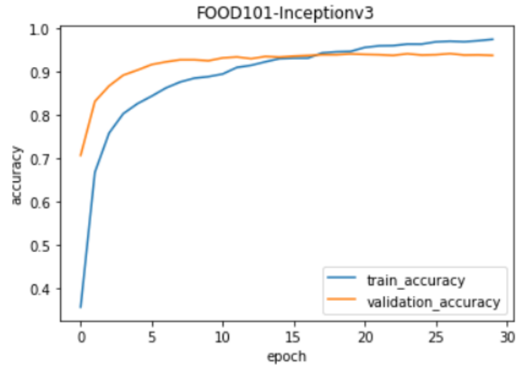

Fig-4: Accuracy vs. Epoch Graph on the dataset with Inception-v3

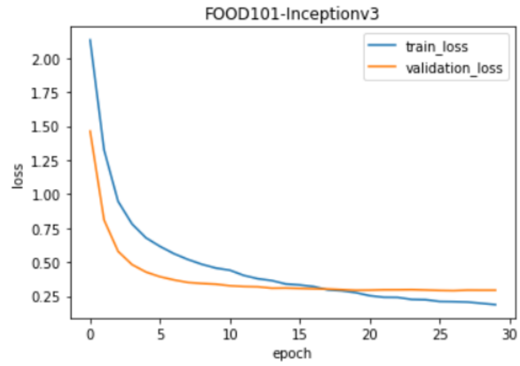

Fig-5: Loss vs. Epoch Graph on the dataset with Inception v-3 Graph

After classifying the image, the nutritional content of that food item was obtained through the API. To use the API, the label of the image was passed which was used to execute the query in the API to obtain the information.

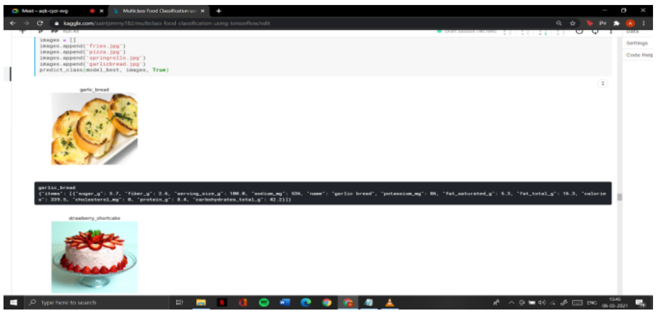

Fig-6: Different food attributes estimated by the proposed algorithm.

5. CONCLUSIONS

Getting an exact value of nutritional value in food items is a complex task since the number of ingredients is not known by only the image given by the user but an estimate of nutritional values can be given as closely as possible. Thus, this analysis addresses the food recognition by exploiting the various acoustic features of the food images. We have used CNN Inception v3 model for learning the high-level features in the image by extracting features from the image. Also, the attributes of the recognized food item were successfully estimated.

In the above result and analysis section we saw our system model performance that is considerably high and as per simulation of the CNN we analyzed that it requires a high performable system for the large number of datasets. The capability of CNN for training is very high for non-linear data, but it requires more “computational time” in order to train the model; though, the performance of the model matters a lot, therefore once the system model is trained properly it takes very less time to produce the system output. As per the performed analysis, we can conclude that our proposed CNNs model is suitable for the food images classification. Our proposed method for estimating attributes also achieved encouraging results. Future endeavors in this domain can include the practical application of this work and more improvements with advanced features to make it a complete guide for everyday meals.

REFERENCES

[1] Raza Yunus, Omar Arif, Hammad Afzal, “To Estimate the Nutritional Value of Food”, December 2018.
[2] Manal Chokr, Shady Elbassouni, “Calories prediction from food images,” in Association for the advancement of artificial intelligence, 2017.
[3] Taichi Joutou and Keiji Yanai, “A food image recognition system with Multiple Kernel Learning,” in 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo,Egypt
[4] Vinay Bettadapura, Edison Thomaz , Aman Parnami, et al,”Leveraging Context to Support Automated Food
[5] Recognition in Restaurants,” in 2015 IEEE Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA.
[6] P. Pouladzadeh, P. Kuhad, S. V. B. Peddi, A. Yassine and S. Shirmohammadi, “ Food calorie measurement using deep learning neural network,” in IEEE International Instrumentation and Measurement Technology Conference Proceedings, Taipei, 2016
[7] Runar Isaksen, Eirik Bø Knudsen, Aline Iyagizeneza Walde “A deep learning segmentation approach to calories estimation of food images,”University of Agder, 2019.
[8] N. Martinel, G. L. Foresti and C. Micheloni, "Wide-Slice Residual Networks for Food Recognition," 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, 2018, pp. 567-576, doi: 10.1109/WACV.2018.00068.
[9] Bossard L., Guillaumin M., Van Gool L. (2014) Food-101 – Mining Discriminative Components with Random Forests. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8694. Springer, Cham. https://doi.org/10.1007/978-3-319-10599-4_29
[10] Viswanath.C.Burkapalli,Priyadarshini.C.Patil,“An Efficient Food Image Classification by InceptionV3 Based Cnns,” in INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 9, ISSUE 03, MARCH 2020.